

La coincidencia entre las notas de prensa y los breves de economía: análisis cuantitativo y cualitativo de dos periódicos españoles

Jesús Martínez

Universitat Autònoma de Barcelona
reporteroriesus@gmail.com

Josep-Lluís Micó

Universitat Ramon Llull
joseplluismis@blanquerna.url.edu

Francesc Pozo

Universitat Politècnica de Catalunya
francesc.pozo@upc.edu



Fecha de presentación: marzo de 2018
Fecha de aceptación: septiembre de 2018
Fecha de publicación: diciembre de 2018

Cita recomendada: MARTÍNEZ, J.; MICÓ, J. L. y POZO, F. (2018). «La coincidencia entre las notas de prensa y los breves de economía: análisis cuantitativo y cualitativo de dos periódicos españoles». *Anàlisi: Quaderns de Comunicació i Cultura*, 59, 29-50. DOI: <<https://doi.org/10.5565/rev/analisi.3143>>

Resumen

La presente investigación tiene como objetivo comparar entre las notas de prensa de empresas y organismos públicos y los breves publicados por los medios de comunicación. Con ello se pretende detectar si la nota de prensa es procesada o contrastada antes de ser publicada o si, por el contrario, es difundida tal como llega a la redacción. Lo que se busca, pues, es establecer numéricamente qué porcentaje del breve es coincidente o similar a la nota de prensa. La muestra ha cubierto medio año, el primer semestre de 2014, y se ha escogido *El Mundo* y *La Vanguardia*, los diarios de información general de pago en España que cuentan con más breves de empresa como parte fija de una sección («Economía»). Posteriormente, una serie de entrevistas en profundidad (entre 2015 y 2018) ha permitido orientar la interpretación de los datos conseguidos. Metodológicamente, una herramienta específica, DetectPlagiarism, ha servido para cotejar textos a través del comando SimilarityScore. Se han obtenido textos tan similares que son susceptibles de haber sido reproducidos igual o de contener alguna porción de texto que ha sido copiada. En el uso del comando DetectPlagiarism, el umbral de copia por defecto es de 0,35. Pues

bien, los índices de similitud entre los breves y las notas de prensa de *La Vanguardia* y *El Mundo* tienen un valor medio de 0,41, superior a este umbral.

Palabras clave: periodismo; géneros; plagio; economía; empresas; comunicación empresarial

Resum. *La coincidència entre les notes de premsa i els breus d'economia: anàlisi quantitativa i qualitativa de dos diaris espanyols*

Aquesta recerca té com a objectiu comparar les notes de premsa d'empreses i organismes públics i els breus publicats pels mitjans de comunicació. Amb això es pretén detectar si la nota de premsa és processada o contrastada abans de ser publicada o si, per contra, és difosa tal com arriba a la redacció. El que es busca, doncs, és establir numèricament quin percentatge del breu és coincident o similar a la nota de premsa. La mostra ha cobert mig any, el primer semestre de 2014, i s'ha escollit *El Mundo* i *La Vanguardia*, els diaris d'informació general de pagament a Espanya que tenen més breus d'empresa com a part fixa d'una secció («Economia»). Posteriorment, una sèrie d'entrevistes en profunditat (entre 2015 i 2018) ha permès orientar la interpretació de les dades aconseguides. Metodològicament, una eina específica, DetectPlagiarism, ha servit per acabar textos mitjançant l'ordre SimilarityScore. S'han obtingut textos tan similars que són susceptibles d'haver estat reproduïts igual o de contenir alguna porció de text que ha estat copiada. Utilitzant l'ordre DetectPlagiarism, el llindar de còpia per defecte és de 0,35. Doncs bé, els índexs de similitud entre els breus i les notes de premsa de *La Vanguardia* i *El Mundo* tenen un valor mitjà de 0,41, superior a aquest llindar.

Paraules clau: periodisme; gèneres; plagio; economia; empreses; comunicació empresarial

Abstract. *Coincidences between press releases and economic briefs: Quantitative and qualitative analysis of two Spanish newspapers*

The objective of this research is to compare the press releases of companies and public organizations and the briefs published by the media. With this goal in mind, it is intended to detect if the press release is processed or contrasted before being published, or if, on the contrary, it is disseminated as it arrives at the newsroom. What is sought, then, is to numerically establish what percentage of the briefs are coincident or similar to the press release. The sample has covered half a year, the first semester of 2014, and *El Mundo* and *La Vanguardia* have been chosen, the general information paid newspapers in Spain that have the largest number of briefs as a part of a regular section (Economy). Subsequently, a series of in-depth interviews (between 2015 and 2018) has made it possible to guide the interpretation of the data obtained. Methodologically, a specific tool, DetectPlagiarism, has served to collate texts through the SimilarityScore command. So similar texts have been obtained that are likely to have been copied or to contain some portion of the text that has been copied. In the use of the DetectPlagiarism command, the default copy threshold is 0.35. In this respect, the indexes of similarity between the briefs and the press releases of *La Vanguardia* and *El Mundo* have an average value of 0.41, above this threshold.

Keywords: journalism; genres; plagiarism; economics; companies; business communication

1. Introducción

Un grupo de estudiantes de Periodismo, dirigidos por la informadora Wendy Bacon, analizaron el contenido de una decena de periódicos australianos durante una semana y llegaron a la conclusión de que más de la mitad de las noticias publicadas (un 55 %) tenían su origen en una nota de prensa. Su investigación, auspiciada por el Australian Center for Independent Journalism y el portal Crikey, les llevó a descubrir que había periodistas en plantilla que firmaban piezas que prácticamente eran la reproducción literal de la información que les habían enviado las empresas e instituciones. De los 2.203 artículos analizados, más de 500 no contenían más puntos de vista, fuentes o contenidos que los aportados por la nota de prensa (Bacon et al., 2010).

Este asunto se ha convertido en una preocupación notable para los investigadores de los medios, en especial en mercados precarios como el español, donde a los redactores se les exige una gran producción con pocos recursos y personal. Así lo atestiguan tesis doctorales como *La fiabilidad del proceso documental del discurso periodístico como fuente de información*, de Inmaculada Chacón (1994); *Estudio de las fuentes de información en el marco del periodismo especializado: Estrategias de selección y tratamiento de las fuentes en las secciones periodísticas de El Mundo y El País*, de Concha Pérez (2005), y *Pragmática de la desinformación: Estrategias e incidencia de la calidad informativa de los medios*, de Luis M. Romero (2014).

Partiendo de los elevados grados de «manipulación» que se dan en la cadena informativa (Durandin, 1982) y digital (Scholz, 2013), otros autores habían llevado a cabo útiles aportaciones sobre el uso y el abuso de la información de agencia en un soporte en el que destacaban la urgencia y la escasez de medios: las ediciones electrónicas de los periódicos (Sancha, 2005).

Por esta razón, hay autores que se refieren a la desinformación que se genera de esta manera (Starr, 2009), puesto que, a su entender, lo que realmente se difunde desde medios que operan sobre todas las plataformas es publicidad, propaganda..., contenidos promocionales, en última instancia, y no información contrastada, elaborada con rigor (Schwoebel, 1971; McChesney, 1999; Manning, 2001). No obstante, hasta ahora no ha habido trabajos centrados exclusivamente en esta práctica tan extendida sobre los breves, un género que, como se verá a continuación, la sufre muy frecuentemente. Ni siquiera los investigadores que apuntaron esta vía la desarrollaron posteriormente (Tellería Roca, 1986; De Fontcuberta, 1993: 16; Parratt, 2008).

El breve no consta en la *Teoría de los géneros periodísticos*, de Gomis (2008). Sin embargo, como es obvio, existe. A menudo se presenta como la antítesis del reportaje, el género «totalizador», la estrella por antonomasia (García Márquez, 2010). No obstante, cumple una función tan importante como cualquier otro formato. Autores como Grijelmo (2006) resaltan que es una noticia que apenas consta de *lead* o entrada, esto es, el «núcleo esencial de la noticia». Así pues, cabe entenderlo no como un resumen, sino como una síntesis de un hecho de interés informativo, con una gran economía de pala-

bras. Su función es dar cuenta inmediata y clara de lo fundamental, e interesar y captar la atención para que el público siga leyendo, viendo o escuchando (Jaraba, 2009).

Generalmente, el breve responde a dos de las 6 W (*how, what, where, who, why* y *when*, en inglés), a saber, *qué* y *quién*. Pero esta no es una norma fija. A veces es necesario que atienda a otras W, sobre todo si se trata del seguimiento de una noticia que ha comenzado a publicarse en días anteriores (Fonnegra, 2010). Sea como fuere, puede escribirse de manera pulcra y elegante. Por ello, Grijelmo ruega que los editores y los redactores estén atentos a los breves, que menudean en los periódicos, para que despejen todas las dudas acerca de la noticia. En cualquier caso, la información de breve —que no es lo mismo que información breve— debe seguir el denominado «estilo informativo», caracterizado por ser «sobrio y escueto, objetivo, en el que no hay sitio para el yo del periodista» (De Fontcuberta, 1993).

El *Libro de redacción* (*La Vanguardia*) (2004) reduce el breve a un subapartado del género noticia (los otros son el reportaje, la entrevista, la crónica, el análisis, la crítica y la tribuna editorial):

Los breves (o *fishes*) son elementos que recogen una noticia de forma corta. Agrupados en un bloque de una o dos columnas, los breves (mínimo dos) presentan una extensión similar y han de contener todos los elementos mínimos para la comprensión de la información. No admiten puntos y aparte. Estas piezas no van datadas, pero, al final, se ha de señalar la autoría, separada del texto por un guion largo entre espacios: si es de un redactor, se firma con iniciales o con inicial y apellido (nunca con el nombre entero), o bien con la palabra *redacción*; si es de agencia, se consigna la agencia correspondiente. (*La Vanguardia*, 2004: 34)

El breve no aparece reseñado independientemente en el *Libro de estilo de El Mundo* (2002), pero sí se explica el contenido de lo que debe ser la noticia «básica»:

La noticia o información básica. Este fue el género más habitual en la prensa diaria, al que se aplican más estrictamente las consideraciones sobre frases y párrafos cortos, entrada directa y desarrollo que puede ser cronológico o piramidal, según la complejidad menor o mayor de los elementos informativos. (*El Mundo*, 2002: 6)

2. Objetivo y metodología

El objetivo de esta investigación es cotejar las notas de prensa redactadas por empresas y entidades públicas y los breves sobre estas organizaciones publicados por los medios de comunicación incluidos en el estudio (*El Mundo* y *La Vanguardia*) en sus páginas de economía. Con esta comparación se pretende detectar si la nota de prensa es procesada o contrastada antes de ser publicada, o si, por el contrario, la nota de prensa es difundida tal y como llega a la

redacción. De este modo, lo que se busca es establecer numéricamente qué porcentaje del breve publicado es coincidente o similar a la nota de prensa.

Aunque no se trata de acusar de plagio a los redactores de los periódicos ni de las agencias de noticias, las herramientas que se usarán sí han sido concebidas para la detección de plagio. En los últimos años, una gran cantidad de programas y aplicaciones han sido creados con este fin, con el interés de reducir y evitar la copia en publicaciones como las científicas (Sánchez-Vega et al., 2013).

Una de las herramientas más extendidas es iThenticate, cuya base de datos comparativa contiene miles de millones de artículos que los editores de más de 530 medios utilizan para detectar problemas de plagio y autoría.

Bibliotècnica (<<http://bibliotecnica.upc.edu>>), la biblioteca digital de la Universitat Politècnica de Catalunya-BarcelonaTech, ha elaborado una extensa lista de programas que pueden encontrarse en la red para detectar el plagio. Además, hay programas que son herramientas de verificación de contenidos para periodistas, como Plagtracker, propuesto por el International Center for Journalists.

En este entorno conviene estar familiarizado con el programa de cálculo Maple, propiedad de Maplesoft, y con el concepto de comando, en relación con un programa o una aplicación informática, cada una de las funciones u órdenes que tienen como finalidad realizar una tarea específica.

De forma más precisa, y en el marco de la metodología de la presente investigación, se describen a continuación algunos de los comandos de Maple utilizados:

- SimilarityScore: coteja el uso de palabras comunes entre dos textos y devuelve un valor o una puntuación entre 0 y 1. Una puntuación de 0 indica que no hay ninguna palabra común entre ambos textos; una puntuación de 1 indica que hay una coincidencia total de palabras.
- JaccardCoefficient: compara dos textos. El resultado es un número entre 0 y 1. Los comandos CosineCoefficient y DiceCoefficient son similares.
- DetectPlagiarism: a partir de dos textos, establece si existe plagio entre ellos.

La muestra cubrió el primer semestre de 2014. Se escogió *El Mundo* y *La Vanguardia* por ser los diarios de información general de pago en España que cuentan con más breves de empresa como parte fija de una sección («Economía»). En cifra de lectores diarios en su categoría, ocupan la segunda y la tercera posición, respectivamente, por detrás de *El País*, que fue descartado por no ofrecer el suficiente volumen ni la regularidad deseada en la publicación de breves de economía (AIMC, 2017: 8). En algunos casos, el mismo breve apareció en los dos rotativos y en un mismo día. Pero la prioridad del estudio no era la coincidencia de breves en los periódicos confrontados, sino el análisis de breves cuya fuente fuese una nota de prensa de empresa, es decir, cuya procedencia fuese una sociedad mercantil, que la envía al diario con fines promocionales.

Para incrementar la probabilidad de llegar a hallazgos que tenían mayor probabilidad de ser representativos de un grupo (Csikszentmihalyi y Larson, 2014: 35-54), la elección de breves fue aleatoria. Su selección siguió el proceso siguiente: se empezó el miércoles 1 de enero de 2014 en los dos periódicos y se siguió con el día siguiente, jueves, pero de la siguiente semana: jueves 9 de enero de 2014. Así, de manera escalada, se había cubierto medio año de los dos rotativos.

En la mayoría de los casos, la autoría de los breves es de agencia. Europa Press es la preferida por *La Vanguardia* y *El Mundo* en sus respectivas secciones de economía, por delante de EFE. También se sigue la fórmula de firma «Agencias» (sin especificar) o «Redacción» (o, en el caso de *El Mundo*, E. M.)

Como ya se ha explicado, existen muchas aplicaciones para la detección de plagio, y la investigación en este campo es extensa y está en crecimiento (Barrón-Cedeño et al., 2013). En esta investigación, el breve se tenía que comparar con la nota de prensa o con el despacho de agencia —fáciles de localizar en la web—, por lo que el problema era menos complejo que en el caso general, sobre literatura científica.

A partir de estas consideraciones, se optó por aplicar la librería EssayTools de Maple para la detección de plagio en el ámbito periodístico. La elección de esta herramienta respondía a las características del estudio —en el que dos textos deben ser comparados— y a la flexibilidad que la programación sobre el entorno Maple permite.

En este contexto, el comando SimilarityScore sirve para comparar el uso de palabras en dos o más textos y devuelve una matriz de puntuaciones para cada par de textos. Una puntuación de 0 indica que no hay ningún tipo de solapamiento entre ambos. De forma contraria, una puntuación de 1 indica que hay un solapamiento total. No obstante, una puntuación de 1 no implica necesariamente que ambos textos sean idénticos.

Por ejemplo:

```
>with(EssayTools):  
>textoA := «Sobre la bella lluvia cayó»:  
>textoB := «Sobre la lluvia bella cayó»:  
>SimilarityScore(textoA, textoB)  
[1]
```

Mientras que:

```
>with(EssayTools):  
>textoC := «París Roma»:  
>textoD := «Londres Barcelona»:  
>SimilarityScore(textoC, textoD)  
[0]
```

Existen, además, varios métodos disponibles para comparar el nivel de similitud entre dos textos, como los llamados CosineCoefficient, JaccardCoefficient, DiceCoefficient y sus equivalentes binarios (Niwattanakul et al., 2013).

Si se consideran, por ejemplo, los textos «El día alto» y «El perro alto» y se elabora una lista con todas las palabras distintas que aparecen en ambos textos, la relación es [el, día, alto, perro]. A continuación, se generan dos vectores que indican cuántas veces aparecen dichas palabras en los textos. El vector para el primer texto es [1 1 1 0], ya que *el* aparece una vez, así como *día* y *alto*, mientras que *perro* no aparece. El vector para el segundo texto es [1 0 1 1]. Con estos dos vectores se debe hacer una operación matemática, que será diferente en función del método usado. En el caso de CosineCoefficient, dados dos vectores, v_1 y v_2 , el coeficiente del coseno se calcula como el producto escalar de ambos vectores entre el producto de sus normas, es decir:

$$\frac{v_1 \cdot v_2}{|v_1| \cdot |v_2|}$$

donde:

$$\begin{aligned} v_1 \cdot v_2 &= \sum_{i=1}^n v_1(i)v_2(i) \in \mathbb{R} \\ |v_1| &= \sqrt{\sum_{i=1}^n v_1(i)^2} \geq 0 \\ |v_2| &= \sqrt{\sum_{i=1}^n v_2(i)^2} \geq 0 \end{aligned}$$

Para los vectores considerados como ejemplo, se tiene:

$$\begin{aligned} v_1 &= [1 \quad 1 \quad 1 \quad 0] \\ v_2 &= [1 \quad 0 \quad 1 \quad 1] \\ v_1 \cdot v_2 &= 2 \\ |v_1| &= \sqrt{3} \\ |v_2| &= \sqrt{3} \end{aligned}$$

Por lo tanto:

$$\frac{v_1 \cdot v_2}{|v_1| \cdot |v_2|} = \frac{2}{3}$$

Cada uno de estos tres coeficientes tiene una versión binaria, en la que no es importante el número de veces que aparece una palabra, sino el hecho de que aparezca o no.

DetectPlagiarism compara textos a través del comando SimilarityScore y devuelve aquellos que superan un determinado umbral de similitud. Retorna los textos que son tan similares que son susceptibles de haber sido copiados o de contener alguna porción de texto que haya sido copiada. Esta es una medida probabilística y no implica necesariamente que dos textos con una puntuación elevada sean, efectivamente, copias el uno del otro o que hayan sido copiados de la misma fuente. En el uso del comando DetectPlagiarism, el umbral por defecto es 0,35 y la métrica de similitud es BinaryJaccardCoefficient. El umbral puede modificarse para alcanzar cualquier valor entre 0 y 1, así como la métrica de similitud utilizada (Wang et al., 2013).

Al comparar dos textos, la longitud de ambos es importante. De forma más precisa, el resultado de la comparación puede verse alterado si se confronta un texto muy breve con otro muy largo. Para solucionar este problema, y partiendo de la base de que, en la presente investigación, el texto más corto es siempre el breve, se hizo una comparación recurrente entre el breve y una parte cada vez más grande de la nota de prensa o del despacho de agencia. Esto se llevó a cabo a través de un algoritmo cuyo resultado es el máximo valor del índice de similitud y el número de palabras de la nota de prensa o despacho de agencia que maximizan esta similitud.

Sobre esta base metodológica, el estudio estadístico se desarrolló sobre una muestra de 52 breves de los diarios *La Vanguardia* y *El Mundo*: 30 piezas de *La Vanguardia* y 22 piezas de *El Mundo*. En adelante, las siglas utilizadas corresponden a *La Vanguardia* (LV), *El Mundo* (EM), nota de prensa (NP) y despacho o nota de agencia (NA). Una casilla coloreada indica la existencia de una nota de prensa, un despacho o una nota de agencia o ambas (tabla 1).

Tras la recopilación de datos (realizada en el primer semestre de 2014), estos fueron ordenados y analizados sistemáticamente. Esta labor se alargó durante el segundo semestre de 2014. Para ampliar el alcance de la investigación (Wolf, 1987; Wimmer y Dominick, 2011), estaba previsto efectuar una veintena de entrevistas semiestructuradas en profundidad. Y así se hizo a lo largo de 2015 y el primer trimestre de 2016, en una primera tanda, y en los cuatro primeros meses de 2018, en una segunda oleada. La interpretación de los participantes en esta segunda fase resultaba fundamental para entender, con todos sus matices, los resultados numéricos obtenidos (Busquet et al., 2006). Este extremo se explica por el perfil de los entrevistados: trabajadores de los medios estudiados, tanto redactores como responsables de área o sección; empleados y directivos de gabinetes y agencias de comunicación, y docentes e investigadores en periodismo con práctica en este campo profesional. En total, una veintena de personas —la mitad mujeres y la otra mitad, hombres— y con diversos grados de experiencia, con una ayudante de comunicación corporativa con menos de diez años de experiencia en un extremo y un directivo de periódico con varias décadas de carrera en el otro.

Tabla 1. Estudio estadístico desarrollado sobre una muestra de 52 breves de *La Vanguardia* (LV) y *El Mundo* (EM). Una casilla coloreada indica la existencia de una nota de prensa, una nota de agencia o ambas

LV	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29
NP	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
NA						X		X		X			X	X	X
	31	33	35	37	39	41	43	45	46	47	48	49	50	51	52
NP	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
NA			X			X	X	X	X	X			X	X	

EM	2	4	6	8	10	12	14	16	18	20	22
NP	X	X	X	X	X	X	X	X	X	X	X
NA			X			X	X	X	X	X	
	31	33	35	37	39	41	43	45	46	47	48
NP	X	X	X	X	X	X	X	X	X	X	X
NA			X			X	X	X	X	X	

Fuente: elaboración propia.

3. Resultados

Se ha comparado, por separado, breve y nota de prensa; breve y despacho de agencia, y nota de prensa y despacho o nota de agencia. En los 30 breves de *La Vanguardia* analizados, 19 se asemejan a las notas de prensa de las multinacionales. Cuando entre el breve y la nota de prensa de la empresa aparece el intermediario, es decir, la agencia de noticias, la similitud también es notable: si en 14 de estos 30 breves de *La Vanguardia* se cuenta también con el despacho de agencia, en 10 casos este despacho de agencia coincide con el breve. Y el despacho coincide plenamente con la nota de prensa original. Cuando se dispone del despacho de agencia y de la nota de prensa correspondiente (14 casos), en 9 ocasiones el breve y la nota de la multinacional son similares.

En cuanto a los 22 breves de *El Mundo* analizados, 12 tienen aspectos coincidentes con la redacción propuesta por la multinacional. En este periódico mayoritariamente interviene el intermediario, esto es, la agencia de noticias (así es en 21 de los 22 casos). De esa manera, el breve mantiene analogías evidentes con la pieza de la agencia. De los 19 casos en los que se dispone del despacho de agencia y de la redacción primigenia, de la multinacional, en 13 ocasiones existen homogeneidades entre ambos.

3.1. *La Vanguardia*

Utilizando el algoritmo de similitud, se han comparado los 30 breves de *La Vanguardia* y las correspondientes notas de prensa. Los resultados, que aparecen con detalle en la tabla 2, se resumen gráficamente en la imagen 1. En las imágenes 1-8 la línea continua representa el umbral de similitud, por encima del cual se puede decir que sí existe similitud, la línea discontinua marca la

Tabla 2. Índice binario de Jaccard, resultado de la comparación entre los breves y las notas de prensa que se indican

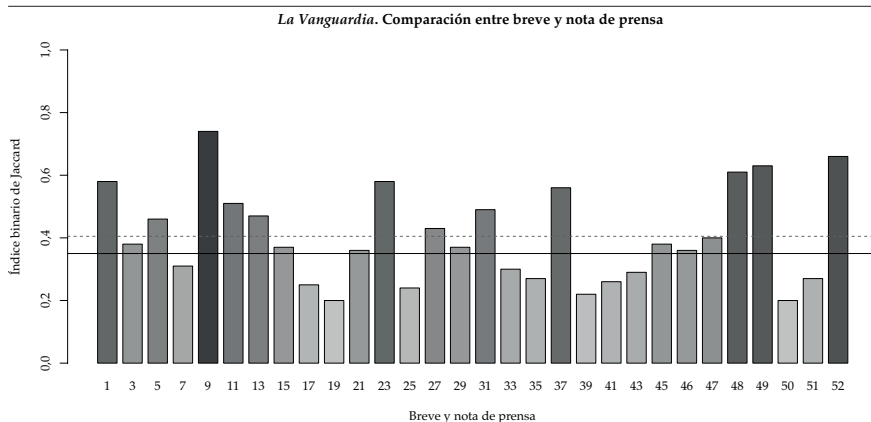
Texto 1	Texto 2	Índice	Palabras	¿Similitud?
B01	NP01	0,58	75	Sí
B03	NP03	0,38	57	Sí
B05	NP05	0,46	133	Sí
B07	NP07	0,31	52	No
B09	NP09	0,74	93	Sí
B11	NP11	0,51	72	Sí
B13	NP13	0,47	57	Sí
B15	NP15	0,37	64	Sí
B17	NP17	0,25	94	No
B19	NP19	0,2	91	No
B21	NP21	0,36	114	Sí
B23	NP23	0,58	82	Sí
B25	NP25	0,24	105	No
B27	NP27	0,43	72	Sí
B29	NP29	0,37	81	Sí
B31	NP31	0,49	50	Sí
B33	NP33	0,3	97	No
B35	NP35	0,27	75	No
B37	NP37	0,56	77	Sí
B39	NP39	0,22	52	No
B41	NP41	0,26	55	No
B43	NP43	0,29	59	No
B45	NP45	0,38	57	Sí
B46	NP46	0,36	86	Sí
B47	NP47	0,4	85	Sí
B48	NP48	0,61	86	Sí
B49	NP49	0,63	86	Sí
B50	NP50	0,2	142	No
B51	NP51	0,27	104	No
B52	NP52	0,66	55	Sí

Fuente: elaboración propia.

media aritmética de los índices de similitud y la intensidad del gris de cada una de las columnas es proporcional al coeficiente de similitud. Del mismo modo, en las tablas 2-7 se incluye la longitud del texto con el que se alcanza la máxima similitud y el resultado final del test.

Cuando el índice binario de Jaccard es superior a 0,35, existe una evidencia estadística de la existencia de una gran similitud. Esto se refleja en la última columna de la tabla 2.

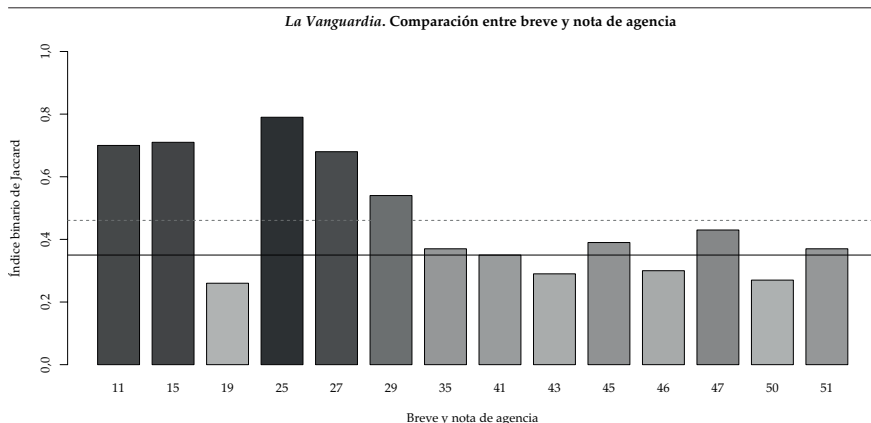
Puede verse cómo, de los 30 pares de textos comparados, en 19 de ellos se obtiene un valor de similitud superior a 0,35, lo que representa el 63,33 % de

Imagen 1. Índice binario de Jaccard al comparar los breves con las notas de prensa

Fuente: elaboración propia.

los textos. Del mismo modo, el valor promedio del índice de similitud se sitúa en 0,41, que es superior a 0,35.

Con el algoritmo de similitud se han comparado los breves de *La Vanguardia* y los correspondientes despachos de agencia. Los resultados, que se resumen en la imagen 2, aparecen con detalle en la tabla 3. Puede observarse que, de los 14 pares de textos comparados, en 10 se obtiene un valor de similitud superior a 0,35, lo que representa el 71,43% de los textos. Además, el valor promedio del índice de similitud se sitúa en 0,46, que es superior a 0,35. En 10 de los breves (de los 14 pares) hay coincidencia.

Imagen 2. Índice binario de Jaccard al comparar los breves con los despachos de agencia

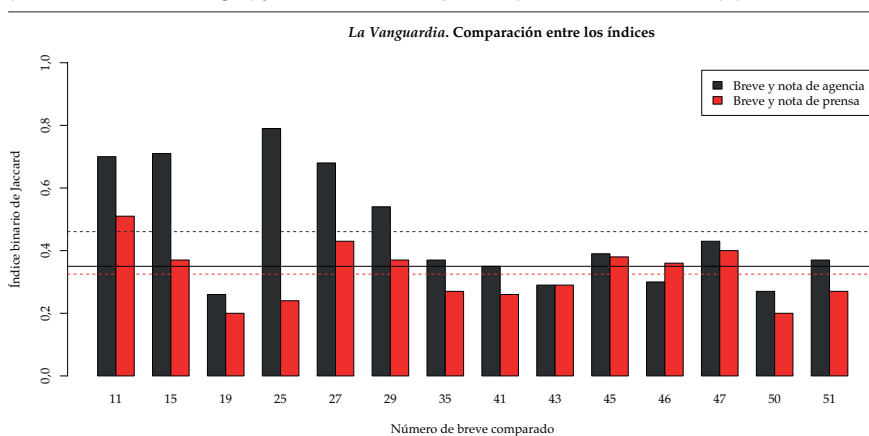
Fuente: elaboración propia.

Tabla 3. Índice binario de Jaccard, resultado de la comparación entre los breves y las notas de agencia que se indican

Texto 1	Texto 2	Índice	Palabras	¿Similitud?
B11	NA11	0,7	63	Sí
B15	NA15	0,71	76	Sí
B19	NA19	0,26	88	No
B25	NA25	0,79	107	Sí
B27	NA27	0,68	96	Sí
B29	NA29	0,54	86	Sí
B35	NA35	0,37	81	Sí
B41	NA41	0,35	53	Sí
B43	NA43	0,29	69	No
B45	NA45	0,39	62	Sí
B46	NA46	0,3	128	No
B47	NA47	0,43	128	Sí
B50	NA50	0,27	144	No
B51	NA51	0,37	85	Sí

Fuente: elaboración propia.

En aquellos casos en los que se dispone tanto de la nota de prensa como del despacho de agencia, se ha establecido también una comparación que se resume gráficamente en la imagen 3. En 14 de los 30 breves de *La Vanguardia* se cuenta con el despacho de agencia y con la fuente original, el comunicado de la empresa. De la observación de dicha ilustración se desprende que el índice de similitud entre breve y despacho de agencia es superior al índice

Imagen 3. Índice binario de Jaccard al comparar los breves con los despachos de agencia (columnas de color negro) y con las notas de prensa (columnas de color rojo)

Fuente: elaboración propia.

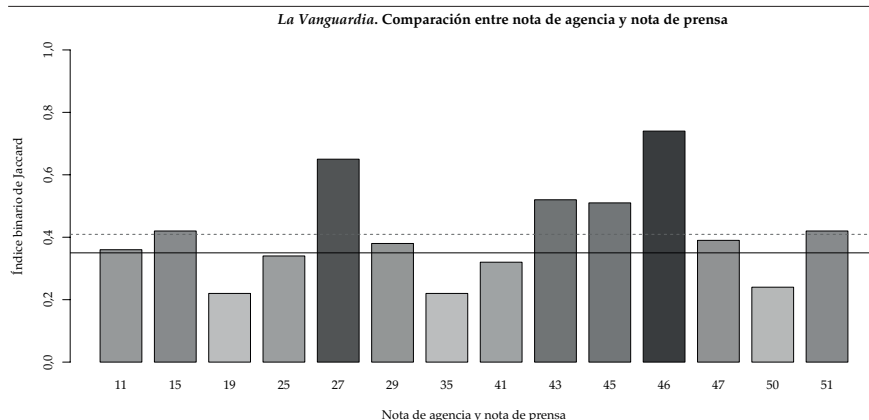
entre breve y nota de prensa en todos los casos salvo en uno. El valor promedio (0,33) de los índices de similitud entre el breve y la nota de prensa —cuando el despacho de la agencia de noticias está también disponible— es inferior al umbral de 0,35. Contrariamente, el promedio de los índices de similitud entre breve y despacho de agencia es, en este caso, de 0,46.

En *La Vanguardia*, cuando se dispone de la nota de prensa y de la nota de agencia, el breve publicado es firmado, en la mayoría de los casos, por una agencia de noticias (Europa Press, EFE o, incluso, como «Agencias»). Con el algoritmo de similitud se ha comparado la afinidad entre las notas de agencia que son fuente de los breves y las correspondientes notas de prensa originales. Los resultados se sintetizan en la imagen 4 y se desglosan en la tabla 4. Puede apreciarse que, de los 14 pares de textos comparados, en 9 se llega a un valor de similitud claramente superior a 0,35, lo que representa el 64,28 % de los textos. El valor promedio del índice de similitud se sitúa en 0,41, superior a 0,35.

3.2. *El Mundo*

Del mismo modo que se ha procedido con *La Vanguardia*, se ha comparado la similitud entre los breves de *El Mundo* y las correspondientes notas de prensa. Los resultados se resumen en la imagen 5 y se muestran con un mayor desarrollo en la tabla 5. Cuando el índice binario de Jaccard es superior a 0,35 se da una evidencia estadística de una gran similitud. Esto se refleja en la última columna de la tabla 5. En 12 de los 20 pares de textos comparados se llega a un valor de similitud superior a 0,35, lo que representa el 60 % de los textos. El valor promedio del índice de similitud se sitúa en 0,41, que es superior a 0,35.

Imagen 4. Índice binario de Jaccard al comparar las notas de prensa con los despachos de agencia



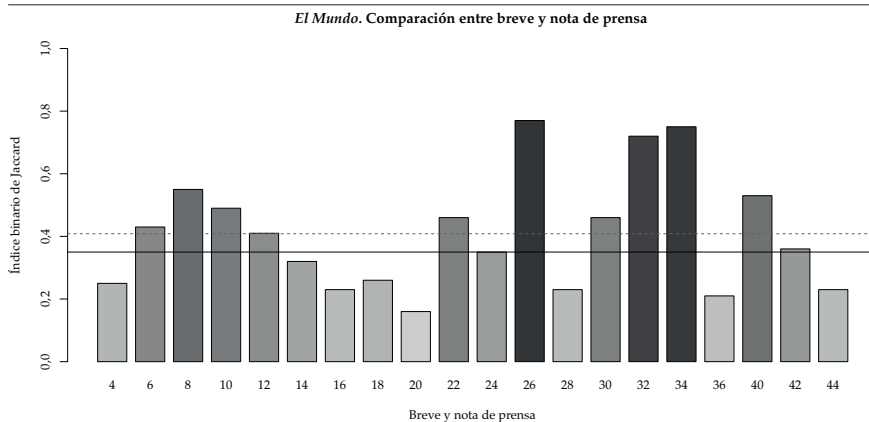
Fuente: elaboración propia.

Tabla 4. Índice binario de Jaccard, resultado de la comparación entre las notas de prensa y las informaciones de agencia que se indican

Texto 1	Texto 2	Índice	Palabras	¿Similitud?
NA11	NP11	0,38	201	Sí
NA15	NP15	0,42	244	Sí
NA19	NP19	0,22	519	No
NA25	NP25	0,34	400	No
NA27	NP27	0,65	318	Sí
NA29	NP29	0,38	140	Sí
NA35	NP35	0,22	219	No
NA41	NP41	0,32	432	No
NA43	NP43	0,52	383	Sí
NA45	NP45	0,51	149	Sí
NA46	NP46	0,74	263	Sí
NA47	NP47	0,39	135	Sí
NA50	NP50	0,24	261	No
NA51	NP51	0,42	282	Sí

Fuente: elaboración propia.

Igualmente, se ha comparado la semejanza entre los breves de *El Mundo* y los correspondientes despachos o notas de agencia de noticias con el algoritmo de similitud. Los resultados, como en las ocasiones anteriores, están en la imagen 6 y la tabla 6. En 20 de los 21 pares cotejados se consigue un valor de similitud superior a 0,35, lo que representa un 95,24 % de los textos. El valor promedio del índice de similitud se sitúa en 0,69, que también es superior a 0,35. La similitud que se establece entre breve y despacho de agencia es superior a la que se da entre breve y nota de prensa.

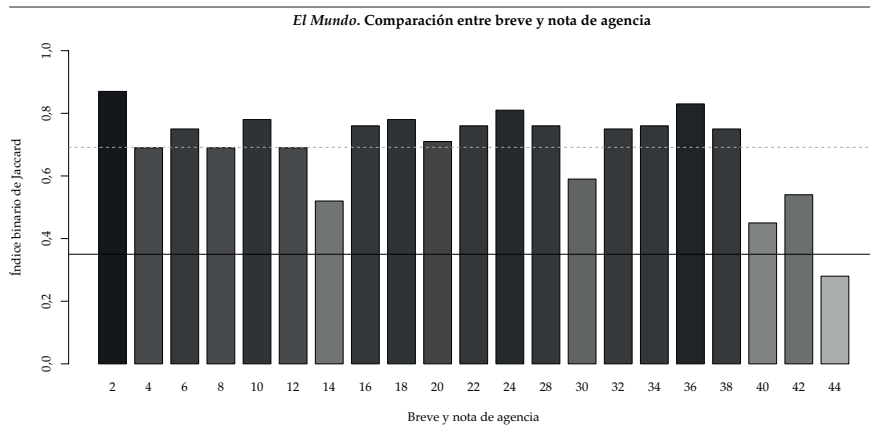
Imagen 5. Índice binario de Jaccard al comparar los breves con las notas de prensa

Fuente: elaboración propia.

Tabla 5. Índice binario de Jaccard, resultado de la comparación entre los breves y las notas de prensa que se indican

Texto 1	Texto 2	Índice	Palabras	¿Similitud?
B04	NP04	0,25	139	No
B06	NP06	0,43	160	Sí
B08	NP08	0,55	102	Sí
B10	NP10	0,49	52	Sí
B12	NP12	0,41	90	Sí
B14	NP14	0,32	104	No
B16	NP16	0,23	59	No
B16	NP18	0,26	44	No
B20	NP20	0,16	36	No
B22	NP22	0,46	149	Sí
B24	NP24	0,35	99	Sí
B26	NP26	0,77	66	Sí
B28	NP28	0,23	127	No
B30	NP30	0,46	57	Sí
B32	NP32	0,72	77	Sí
B34	NP34	0,75	95	Sí
B36	NP36	0,21	48	No
B40	NP40	0,53	74	Sí
B42	NP42	0,36	61	Sí
B44	NP44	0,23	64	No

Fuente: elaboración propia.

Imagen 6. Índice binario de Jaccard al comparar los breves con las notas de agencia

Fuente: elaboración propia.

Tabla 6. Índice binario de Jaccard, resultado de la comparación entre los breves y los despachos de agencia que se indican

Texto 1	Texto 2	índice	Palabras	¿Similitud?
B04	NA04	0,87	52	Sí
B06	NA06	0,69	130	Sí
B08	NA08	0,75	117	Sí
B10	NA10	0,69	78	Sí
B12	NA12	0,78	60	Sí
B14	NA14	0,69	92	Sí
B16	NA16	0,52	108	Sí
B18	NA18	0,76	44	Sí
B20	NA20	0,78	46	Sí
B22	NA22	0,71	36	Sí
B24	NA24	0,76	101	Sí
B26	NA26	0,81	121	Sí
B28	NA28	0,76	165	Sí
B30	NA30	0,59	75	Sí
B32	NA32	0,75	77	Sí
B34	NA34	0,76	101	Sí
B36	NA36	0,83	48	Sí
B38	NA38	0,75	56	Sí
B40	NA40	0,45	70	Sí
B42	NA42	0,54	56	Sí
B44	NA44	0,28	65	No

Fuente: elaboración propia.

Cuando se tenía tanto la nota de prensa como el despacho de agencia de noticias se ha establecido una comparación que se muestra en la imagen 7. El índice de similitud entre breve y despacho de agencia es superior al índice entre breve y nota de prensa en todos los casos, a excepción de uno.

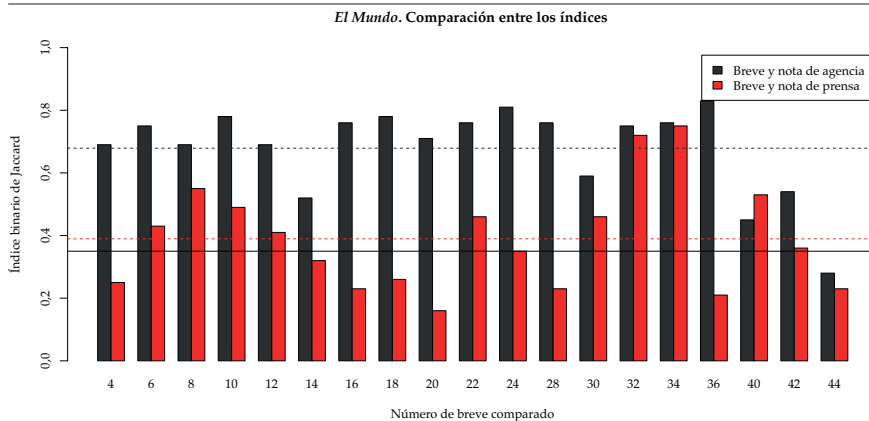
En el caso del diario *El Mundo* todos los breves analizados en esta investigación, a excepción de uno, están firmados por agencias de noticias: EFE y, especialmente, Europa Press.

El algoritmo de similitud se ha empleado para estudiar la relación entre los despachos de agencia de los que beben los breves de *El Mundo* y las correspondientes notas de prensa originales. Los resultados aparecen en la imagen 8 y la tabla 7. Para 13 de los 19 pares de textos comparados, el valor de similitud es superior a 0,35, lo que representa un 68,42 % del total. El promedio del índice de similitud se sitúa en 0,47, de nuevo superior a 0,35.

4. Conclusiones

A continuación se comprobará si existen diferencias significativas entre ambos medios en términos de índices de similitud. En la tabla 8 se ha incluido un

Imagen 7. Índice binario de Jaccard al comparar los breves con los despachos de agencia (columnas de color negro) y con las notas de prensa (columnas de color rojo)



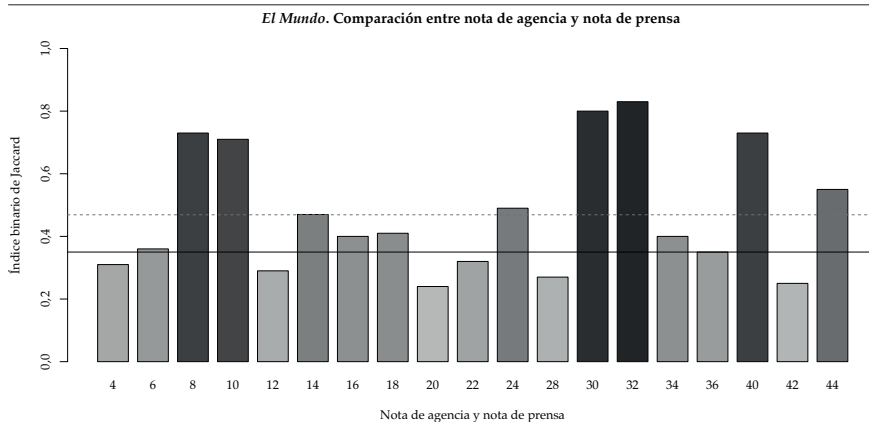
Fuente: elaboración propia.

resumen que incluye la media aritmética y la desviación típica para los índices de similitud entre breve y nota de prensa (NP), y entre breve y nota de agencia (NA) en ambos periódicos.

Los índices de similitud por separado pueden verse en las imágenes 9 y 10, diseñadas para las notas de prensa y los despachos de agencia, respectivamente.

Estas representaciones visuales evidencian que los índices de similitud entre los breves y las notas de prensa de *La Vanguardia* y *El Mundo* están próximos, aunque en *El Mundo* la dispersión de los datos es algo mayor. Sí se

Imagen 8. Índice binario de Jaccard al comparar las notas de prensa con los despachos de agencia



Fuente: elaboración propia.

Tabla 7. Índice binario de Jaccard, resultado de la comparación entre las notas de prensa y los despachos de agencia que se indican

Texto 1	Texto 2	Índice	Palabras	¿Similitud?
NA04	NP04	0,31	681	No
NA06	NP06	0,36	276	Sí
NA08	NP08	0,73	166	Sí
NA10	NP10	0,71	345	Sí
NA12	NP12	0,29	165	No
NA14	NP14	0,47	615	Sí
NA16	NP16	0,4	484	Sí
NA16	NP18	0,41	339	Sí
NA20	NP20	0,24	266	No
NA22	NP22	0,32	276	No
NA24	NP24	0,49	458	Sí
NA28	NP28	0,27	305	No
NA30	NP30	0,8	324	Sí
NA32	NP32	0,83	162	Sí
NA34	NP34	0,4	248	Sí
NA36	NP36	0,35	178	Sí
NA40	NP40	0,73	185	Sí
NA42	NP42	0,25	207	No
NA44	NP44	0,55	223	Sí

Fuente: elaboración propia.

observa una diferencia notable en los índices de similitud entre los breves y los despachos de agencia. El valor promedio es de 0,46 en *La Vanguardia*, mientras que el valor aumenta hasta 0,69 en *El Mundo*. Se llevará a cabo un contraste de hipótesis para confirmar o descartar la intuición derivada de esta diferencia.

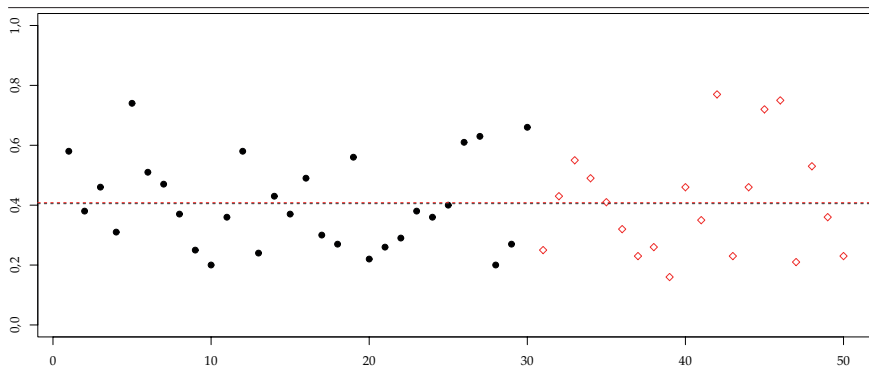
El contraste de hipótesis permite, de forma rigurosa y según criterios científicos, saber si los índices de similitud entre breve y nota de agencia en *La Vanguardia* en relación con los índices de similitud entre breve y nota de agencia en *El Mundo* son significativamente distintos. El contraste de hipótesis se ha basado en el llamado *test de Welch-Satterthwaite*, que puede encontrarse resumido en el trabajo de Mujica et al. (2014). La hipótesis nula (H0)

Tabla 8. Resumen de las estadísticas más comunes (media y desviación típica) para los índices de similitud entre breve y nota de prensa, y entre breve y despacho de agencia, para *La Vanguardia* y *El Mundo*

	<i>La Vanguardia</i>		<i>El Mundo</i>	
	Media	Desviación típica	Media	Desviación típica
NP	0,41	0,15	0,41	0,18
NA	0,46	0,19	0,69	0,14

Fuente: elaboración propia.

Imagen 9. Índices de similitud entre breve y nota de prensa en *La Vanguardia* (puntos negros) y *El Mundo* (rombos rojos). Las líneas discontinuas negra y roja (en la práctica, superpuestas) representan los valores promedios en cada caso

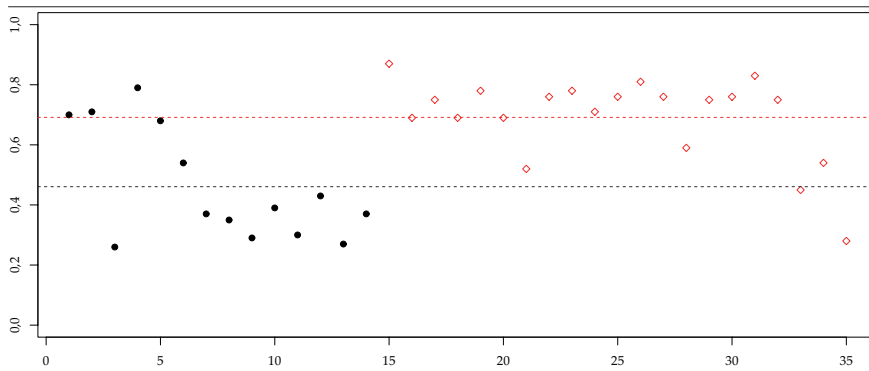


Fuente: elaboración propia.

era que ambos conjuntos no presentaban diferencias con relación a su distribución, mientras que la hipótesis alternativa (H1) establecía que ambos conjuntos se distribuían de forma diferente. Para el contraste de hipótesis, se estableció un nivel de significación del 10 % y se empleó el programa libre R.

Para el conjunto de datos comentado, el p -valor del test es igual a 0,069 %, que es claramente inferior al 10 % del nivel de significación, lo que indica que las diferencias entre ambos conjuntos de datos son estadística y significativamente diferentes. Por ello se puede inferir que en *El Mundo* las similitudes entre breves y despachos de agencia son mayores que en *La Vanguardia*.

Imagen 10. Índices de similitud entre breve y despacho de agencia en *La Vanguardia* (puntos negros) y *El Mundo* (rombos rojos). Las líneas discontinuas negra y roja representan los valores promedios en cada caso



Fuente: elaboración propia.

En 20 de los 52 breves estudiados la firma que aparece es «Redacción»; 35 de los 52 breves reproducen despachos de agencia, y 32 de estos breves sí atribuyen la fuente a las agencias (*El Mundo* aventaja a *La Vanguardia* en este aspecto: solo 1 de los 22 breves estudiados está firmado por la redacción). Así pues, la mayoría de las firmas de los breves son de agencia (32). En estos casos, la agencia de noticias ha copiado mayoritariamente la nota de prensa original. Existe similitud con la nota de prensa en 21 de estos 32 breves supuestamente elaborados por las agencias de noticias (Europa Press, mayoritariamente). En el resto de casos (20), de los 52 breves, su firma corresponde a la redacción.

En 14 de los 19 breves firmados por redactores de la sección de economía de *La Vanguardia* existe similitud con la nota de prensa original. Y 3 de estos 14 breves son en realidad de agencia. En lo que concierne a *El Mundo*, solo 1 de los breves estudiados está firmado desde la redacción. No obstante, esta pieza guarda enormes similitudes con la nota de prensa original.

La interpretación de estos datos que hacen los entrevistados es coincidente. El sector se ha familiarizado con este *modus operandi* y apenas nadie se sorprende ya en la industria de los medios de que se trabaje de esta manera. Esta mecánica es rápida y resulta barata: requiere poca mano de obra y esta, además, no debe tener una gran cualificación para proceder de esta forma.

Como ha sucedido históricamente (McCombs y Shaw, 1972; Manning, 2001), las empresas exigen a sus gabinetes y departamentos de comunicación que les aseguren una cierta presencia en los medios. Por esta razón, sus profesionales han aprendido a presentar la información para que los periodistas creen que ese material no requiere ser reelaborado, según recuerdan los entrevistados. Y añaden que las agencias tienen prisa por ser las primeras en distribuir cualquier novedad, lo que supuestamente las invita a copiar y pegar las notas de prensa. Cuando sus despachos llegan a las redacciones, como subrayan los participantes en esta investigación, se reproduce un comportamiento similar por parte de los periodistas. Así, concluyen, la nota de prensa, a pesar de haber pasado por dos redacciones (la de la agencia y la del medio), prácticamente no experimenta ninguna alteración.

Referencias bibliográficas

- AIMC (2017). *EGM. Resumen general. Febrero a noviembre 2017*. Madrid: Asociación para la Investigación de Medios de Comunicación.
- BACON, W.; LOH, M.; TAYLOR, A. y PAVEY, S. (2010). *Spinning the Media: Key findings in a week in the life of the media* [en línea]. <<https://www.crikey.com.au/2010/03/15/spinning-the-media-key-findings-in-a-week-in-the-life-of-the-media/>> [Fecha de consulta: 05/03/2018].
- BARRÓN-CEDENO, A.; GUPTA, P. y ROSSO, P. (2013). «Methods for cross-language plagiarism detection». *Knowledge-Based Systems*, 50, 211-217. <<https://doi.org/10.1016/j.knosys.2013.06.018>>
- BUSQUET, J.; MEDINA, A. y SORT, J. (2006). *La recerca en comunicació: Què hem de saber? Quins passos hem de seguir?* Barcelona: Editorial UOC.

- CHACÓN, I. (1998). *La fiabilidad del proceso documental del discurso periodístico como fuente de información*. Madrid: Universidad Complutense de Madrid. Tesis doctoral.
- CSIKSZENTMIHALYI, M. y LARSON, R. (2014). «Validity and Reliability of the Experience-Sampling Method». En: CSIKSZENTMIHALYI, M. *Flow and the Foundations of Positive Psychology*. Nueva York: Springer, 35-54.
- DURANDIN, G. (1982). *La mentira en la propaganda política y en la publicidad*. Barcelona: Paidós.
- FONTCUBERTA, M. de (1993). *La noticia: Pista para percibir el mundo*. Barcelona: Paidós.
- FONNEGRA, N. M. (2010). «Lead y estructura de la noticia». En el blog *Nazly María Fonnegra*, el 23 de febrero del 2010 [en línea]. <<http://nazlymariafonnegra.blogspot.com/2010/02/lead-y-estructura-de-la-noticia.html>> [Consulta: 1 de noviembre del 2014].
- GARCÍA MÁRQUEZ, G. (2010). *Yo no vengo a decir un discurso*. Barcelona: DeBolsillo.
- GOMIS, L. (2008). *Teoría de los géneros periodísticos*. Barcelona: Editorial UOC.
- GRIJELMO, Á. (2006). *El estilo del periodista*. Madrid: Taurus.
- JARABA, G. (2009). *Estructura de la noticia: El lead y el cuerpo*. *Periodismo UAB wiki* [en línea]. <<http://periodismouab.wikispaces.com/space/content?o=20>> [Fecha de consulta: 05/07/2016].
- Libro de estilo de El Mundo*. Sevilla: Comunicación y Proyectos Editoriales en Contexto, 2002.
- Libro de redacción (La Vanguardia)*. Barcelona: Ariel, 2004.
- MANNING, P. (2001). *News and news sources: A critical introduction*. Londres: Sage.
- MCCHESENEY, R. W. (1999). *Rich media, poor democracy: Communication politics in dubious times*. Chicago: University of Illinois Press.
- MCCOMBS, M. E. y SHAW, D. L. (1972). «The agenda-setting function of mass media». *Public Opinion Quarterly*, 36 (2), 176-187.
- MUJICA, L. E.; RUIZ, M.; POZO, F.; RODELLAR, J. y GÜEMES, A. (2014). «A structural damage detection indicator based on principal component analysis and statistical hypothesis testing». *Smart Materials and Structures*, 23 (2). <<https://doi.org/10.1088/0964-1726/23/2/025014>>
- NIWATTANAKUL, S.; SINGTHONGCHAI, J.; NAENUDORN, E. y WANAPU, S. (2013). «Using of Jaccard Coefficient for Keywords Similarity». *Proceedings of the International MultiConference on Engineers and Computer Scientists*. Hong Kong (China).
- PARRATT, S. F. (2008). *Géneros periodísticos en prensa*. Quito: Ediciones Ciespal.
- PÉREZ, C. (2005). *Estrategias de selección y tratamiento de las fuentes en las secciones periodísticas de El Mundo y El País*. Sevilla: Universidad de Sevilla. Tesis doctoral.
- ROMERO, L. M. (2014). *Pragmática de la desinformación: Estrategias e incidencia de la calidad informativa de los medios*. Huelva: Universidad de Huelva. Tesis doctoral.
- SANCHA, D. (2005). «El uso de la información de agencia en las ediciones electrónicas de diarios en España». *Athenea Digital*, 8. <<https://doi.org/10.5565/rev/athenead/v1n8.231>>.
- SÁNCHEZ-VEGA, F.; VILLATORO-TELLO, E.; MONTES-Y-GÓMEZ, M.; VILLASEÑOR-PINEDA, L. y ROSSO, P. (2013). «Determining and characterizing the reused text for plagiarism detection». *Expert Systems with Applications*, 40, 1804-1813.
- SCHOLZ, T. (2013). *Digital labor: The Internet as playground and factory*. Nueva York: Routledge.

- SCHWOEBEL, J. (1971). *La prensa, el poder y el dinero*. Barcelona: Dopesa.
- TELLERÍA ROCA, E. (1986). *Diccionario periodístico*. Santiago de Cuba: Editorial Oriente.
- WANG, S.; QI, H.; KONG, L. y NU, C. (2013). «Combination of VSM and Jaccard coefficient for external plagiarism detection». *Proceedings of the 2013 International Conference on Machine Learning and Cybernetics*. Tianjin (China).
- WEILL, G. (2007). *El periódico: Orígenes, evolución y función de la prensa periódica*. Sevilla: Comunicación Social.
- WIMMER, R. y DOMINICK, J. (2011). *Mass media research: An introduction*. Boston: Cengage Learning.
- WOLF, M. (1987). *La investigación de la comunicación de masas: Crítica y perspectivas*. Barcelona: Paidós.